APPENDIX D

Spatial Autoregression Analysis of the PCMe Data

Daniel A. Griffith, Department of Geography, University of Miami

Test Only Data

Description of sample. A total of 4,316 samples were collected, of which 10 have no geographic labels. The total number of samples with a value that exceeds the threshold level is 21. These sample asbestos measurements were aggregated by location into 45 statistical summary areas (SSAs) for lower Manhattan (south of Canal Street). One of these SSAs is the site that housed the WTC; no data were collected for this plus an additional 8 SSAs.

Initial data analysis. The rareness of exceedances suggests that these data may be described by a Poisson model. One feature of a Poisson random variable is that its mean, μ, and its variance are equal (equidispersion), a property frequently violated by real world data. "Failure of the Poisson assumption of equidispersion has similar qualitative consequences to failure of the assumption of homoskedasticity" associated with the Gaussian distribution (Cameron and Trivedi, 1998, p. 77). The standard way of accommodating overdispersion (the presence of more variation than is expected for a Poisson random variable) is by replacing a Poisson random variable with a negative binomial random variable—which can be viewed as a gamma mixture of Poisson random variables. In doing so, the distribution of counts is viewed as either (1) having missing variables for the mean specification, or (2) being dependent (i.e., the occurrence of an event increases the probability of further events occurring). The most popular implementation of the negative binomial probability model specifies the variance as being quadratic in the mean, or

$$\mu + \eta\mu^2 = (1 + \eta\mu)\mu,$$

with the dispersion parameter, η, to be estimated. The magnitude of η may be interpreted as follows (after Cameron and Trivedi, 1998, p. 79):

$\eta = 0$ implies no overdispersion;

$\eta \approx \frac{1}{\mu}$ implies a modest degree of overdispersion; and,

$\eta \geq \frac{2}{\mu}$ implies considerable overdispersion.

In other words, if $0 \leq \eta < \frac{0.5}{\mu}$, a spatial analyst may consider overdispersion detected in georeferenced data to be inconsequential, with little to be gained by replacing a Poisson with a negative binomial model specification. Meanwhile, recognizing that these exceedances are constrained by the number of samples collected suggests that these data may be described by a binomial model. Recoding counts of

exceedances to a binary (0-1) presence/absence measurement suggests that these data may be described by a logistic model. Simple estimation results for each of these four models appear in Table D-1.

Table D-1. Selected model estimation results.

Model	intercept	equidispersion
Poisson	-5.3232	NA
Negative binomial	-5.0964	4.6066
Binomial	-5.3183	NA
Logistic	1.4213	NA

One important implication from the tabulated results appearing in Table D-1 is that a Poisson model description of rates may suffer from a marked violation of the equidispersion assumption. The following evidence supports this claim:

$$\frac{2}{\hat{\mu}} = \frac{2}{0.58333} \approx 3.42857 < 4.6066.$$

In other words, the mean and variance may not be constant across the 36 SSAs.

Accounting for spatial autocorrelation. A conventional spatial autocorrelation analysis is hindered by two features of the collected data. One is the rareness of exceedance. In order to further explore spatial dependency in this context, average measures of asbestos also were analyzed. The other drawback is the absence of data for 9 SSAs. Because these areal units are dispersed across the study region, computing a Moran Coefficient (MC) becomes problematic.

MC scatterplots appear in Figure D-1. No conspicuous geographic pattern is apparent for either rates or averages, in part because of the presence of a large number of zeroes.

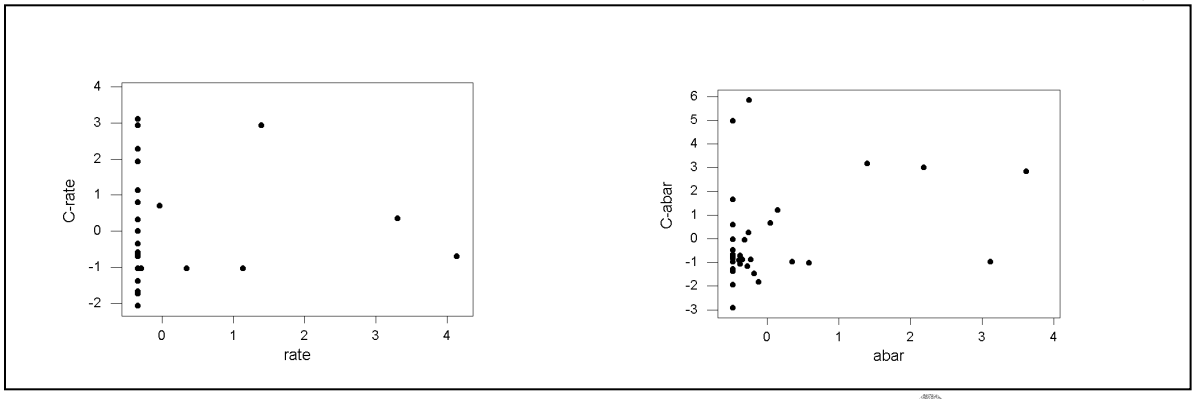

Figure D-1. Left: MC scatterplot for the rate of exceedance. Right: MC scatterplot for the average measure of asbestos.

Latent map patterns also can be assessed with eigenvectors derivable from a MC. Here four of the 11 eigenvectors (E_1, E_8, E_{10}, E_{22}; these were selected using the stepwise options for PROC LOGISTIC in SAS, and SWPOIS in STATA) denoting consequential positive spatial autocorrelation help describe the geographic distribution of exceedance rates. Maps of these synthetic geographic variables appear in Figure D-2.

Figure D-2. Geographic distribution of relevant eigenvectors. Top left: E_1. Top right: E_8. Bottom left: E_{10}. Bottom right: E_{22}.

The MC values for the four eigenvectors range from 0.37 to 0.97. Estimation results that include these eigenvectors in the specifications of each of the four models appearing in Table D-1 are reported in Table D-2. The Poisson model with an assumption of equidispersion appears to be reasonable here. This specification accounts for nearly 60% of the variation in the geographic distribution of rates. The binomial model specification accounts for about 30%. The logistic model description seems inappropriate.

Table D-2. Selected model estimation results when spatial dependence is included.

Model	intercept	equidispersion	E_1	E_8	E_{10}	E_{22}
Poisson	-6.0625	NA	4.6209	-9.2072	3.9665	NA
Negative binomial	-6.1506	0.4476	3.8439	-8.9830	4.5748	NA
Binomial	-6.0572	NA	4.6680	-9.3192	3.9896	NA
Logistic	2.0052	NA	NA	NA	-9.3199	-6.6709

One important finding that can be gleaned from Table D-2 is detected overdispersion accompanying the simple Poisson model description principally is attributable to latent spatial autocorrelation ($0.4476 << \frac{0.5}{0.58333} \approx 0.8571$). Accordingly, these data can be well described with a Poisson model when the model specification captures spatial dependencies.

Pairwise comparisons between SSAs. Pairwise comparisons of SSA asbestos exceedance sampling results were made to assess whether or not statistically significant differences exist. Aggregate sample sizes less than 30 are considered too unreliable, and were not included in this assessment. The outcome of this sample size restriction is 22 SSAs with a sufficient number of samples, allowing ($22 \times 21/2 =$) 231 pairwise comparisons.

Pairwise comparisons statistical theory. The rareness of exceedances suggests that an analysis of differences of means cannot easily be based upon a binomial model. For a normal approximation to be reasonable here, each SSA sample would need to satisfy the constraint of

(sample size) × (exceedance rate) > 5.

Although both the binomial and the Poisson regression models produce very similar descriptive results for the lower Manhattan asbestos data, the Poisson model seems to furnish a better model-based inferential basis.

Consider the difference between two Poisson random variables with means μ_1 and μ_2. Mathematical statistical theory states that the expected value of the difference of any two random variables equals the difference of their expected values. Therefore, the difference of means for two Poisson random variables equals $\mu_1 - \mu_2$. If these two Poisson random variables are independent, then their difference has a known statistical distribution (Skellam, 1946). The respective sampling variance of each is $\frac{\mu_1}{n_1}$ and $\frac{\mu_2}{n_2}$; the sampling variance of their difference is $\frac{\mu_1}{n_1} + \frac{\mu_2}{n_2}$, which parallels a standard result for normal curve theory. As the two means, μ_1 and μ_2, increase to infinity, the distribution of the difference of these two independent Poisson variables rapidly converges to normality. Convergence on a normal probability distribution is quick, with a very good approximation attained once μ_1>4 and μ_2>4. But for small values of μ_1 and/or μ_2 this normal approximation is poor. In these latter cases, the difference of two Poisson random variables still tends to conform to a Poisson distribution.

When multiple comparisons are being made, the overall level of significance often should be adjusted downward to compensate for an increase in chance null hypothesis rejections (i.e., Type I errors). For example, in the single WTC asbestos study for which 231 difference of means null hypotheses are being evaluated, each hypothesis with a single test, setting the global Type I error probability at α=0.05 means that at least one in twenty of the hypotheses tested will turn up significant, merely due to chance fluctuation. In other words, there is a very good chance of finding at least one test (and as many as 11 or 12) to be statistically significant solely due to sampling variability, incorrectly concluding that a difference exists in the population. The Bonferroni correction/adjustment is the most basic procedure for modifying α to compensate for this increase in Type I error probability. When the samples are independent, the modification becomes $\frac{\alpha}{\#\ \text{of tests}}$. For the WTC study, and a two-tailed test, this becomes $\frac{0.005}{231}$ for an overall α=0.01, $\frac{0.025}{231}$ for an overall α=0.05, and $\frac{0.05}{231}$ for an overall α=0.10. As correlation between the samples increases, the denominator of this adjustment effectively decreases toward 1. Uncorrelated variables require a full Bonferroni adjustment, perfectly correlated variables require no adjustment, and partially correlated variables required an adjustment between these two extremes.

Differences of exceedance rates. The estimated spatially filtered Poisson model produces sample mean estimates for uncorrelated Poisson variables. These models include LN (# of cases) as an offset variable. Therefore, dividing both sides of the estimated equation for $\hat{\mu}_i$ (i.e., the mean rate for areal unit i) by the

corresponding number of samples yields the set of estimated rates, assuming an underlying Poisson process, of $\frac{\hat{\mu}_i}{n_i}$, i=1, 2, …, 22. The accompanying set of null hypotheses becomes

$$H_o: \frac{\mu_i}{n_i} - \frac{\mu_j}{n_j} = 0 \text{ , } i \neq j, \; i=1, 2, \ldots, 22 \text{ and } j=i+1, i+2, \ldots, 22 \text{ .}$$

The estimated standard error for this difference of rates test is given by $\sqrt{\frac{\hat{\mu}_i}{n_i^2} + \frac{\hat{\mu}_j}{n_j^2}}$.

A simulation experiment involving 50,000 difference of means replications (total=231 × 50,000) was conducted using the spatially filtered Poisson model estimation results. The simulated Poisson random variable, Y, then was used in a bivariate linear regression analysis, which yielded

$$\frac{\hat{\mu}_i}{n_i} - \frac{\hat{\mu}_j}{n_j} = -0.00000 + 1.00012 \overline{\frac{y_i}{n_i} - \frac{y_j}{n_j}} + e \text{ , } R^2 = 1.00, \text{ and}$$

$$\sqrt{\frac{\hat{\mu}_i}{n_i^2} + \frac{\hat{\mu}_j}{n_j^2}} = -0.00000 + 1.00031 \, s_{\frac{y_i}{n_i} - \frac{y_j}{n_j}} + e, \; R^2 = 1.00.$$

In both cases, the intercept is not significantly different from 0, and the slope is not significantly different from 1. These simulations confirmed the preceding theoretical results.

The model-based mean estimates range from roughly 0.01 to 7.22, implying that most all of the difference of rates sampling distributions should be non-normal. Each simulated dataset was subjected to a diagnostic Kolmogorov-Smirnov goodness-of-fit test for a normal distribution, producing test statistics ranging from roughly 0.48 to 0.53. In other words, the simulated sampling distributions fail to conform to normal distributions. Consequently, the pairwise difference of rates assessments are based upon a Hope-type nonparametric simulation analysis, involving 99,999 replications coupled with each observed difference. The simulated distribution is based on a pair of Poisson random variables, each with the same mean of $\frac{n_1\mu_2 + n_2\mu_1}{2n_1n_2}$, which yields a null hypothesis difference of 0 and the correct theoretical variance of $\frac{\mu_1}{n_1} + \frac{\mu_2}{n_2}$. Because a two-tailed test is employed here, an observed rank of 1-2 or 99,999-100,000

results in a rejection of the null hypothesis for α=0.01, an observed rank of 3-12 or 99,990-99,998 results in a rejection of the null hypothesis for α=0.05, and an observed rank of 12-22 or 99,979-99,989 results in a rejection of the null hypothesis for α=0.10. Based on these criteria, 21 pairs of exceedance rates are significantly different at the 10% level, 17 pairs are significantly different at the 5% level, and 48 pairs are significantly different at the 1% level. Basically, roughly 37% of the extreme MCGB mean pairs tend to be significantly different. These differences arise from four clusters of mean sizes. The first is dominated by the largest MCBG mean of roughly 9 (MCBG 10015022). The second is dominated by the second and third largest means of approximately 2-3 (MCBG 10008002, MCBG 10015021). The third is dominated by the medium mean of approximately 1.5 (MCBG 10015012). The fourth group is dominated by the relatively small mean of roughly 0.1 (MCBG 10317019D). Significant pairwise contrasts appear in Tables D-3a and D-3b, and Figure D-3.

These results need to be moderated by keeping in mind that the estimated Poisson model accounts for only about 50% of the variance in the observed exceedances.

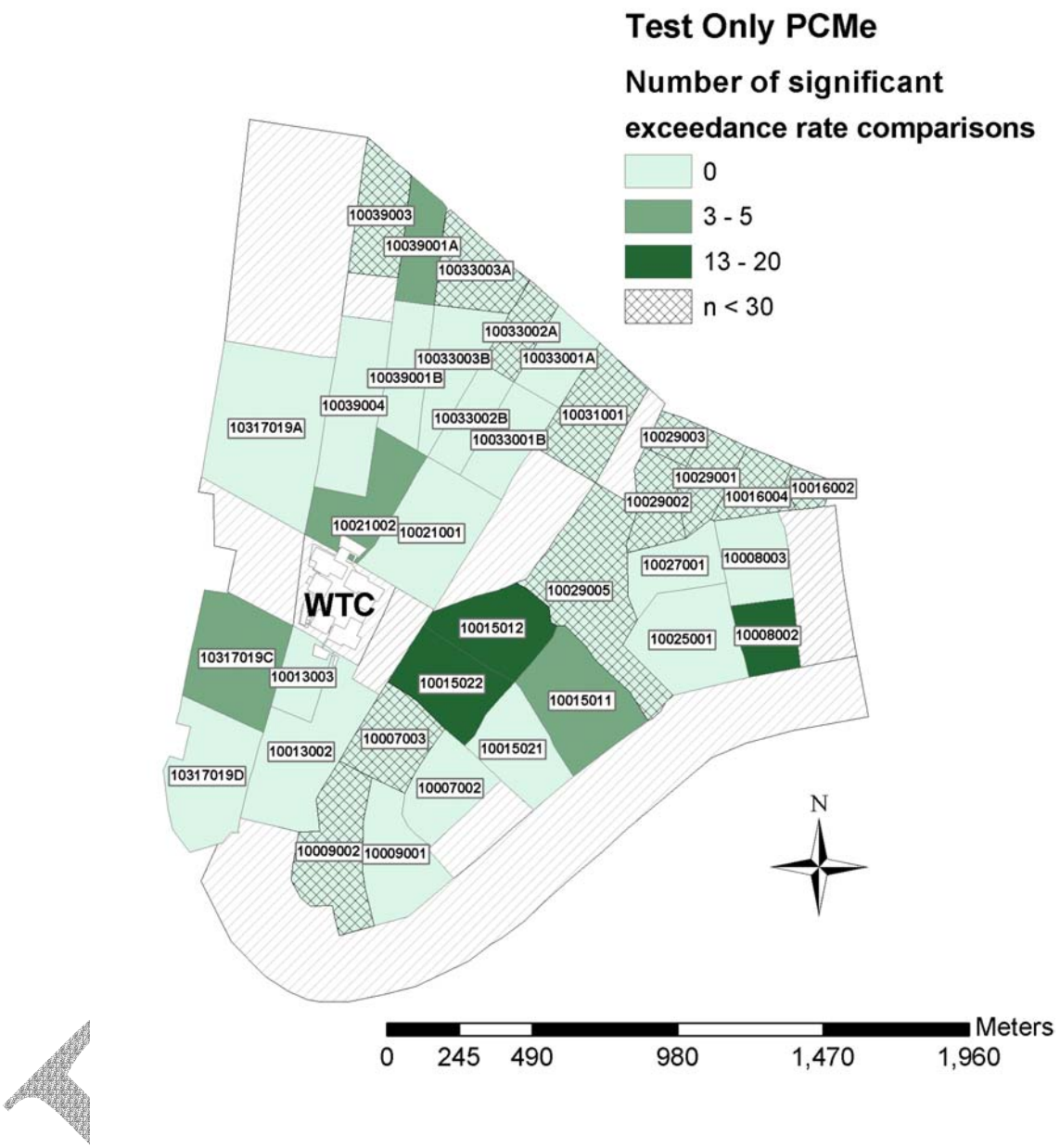

Figure D-3. Significant differences between estimated exceedance rates for *test only* data, with Statistical Summary Areas labeled. Estimates are based on the spatially-filtered Poisson model (see Section 3.2.3.2 and Appendix D for details). The number of significant pairwise comparisons at an experiment-wise $\alpha = 0.01$ (with a Bonferroni adjustment) are shown for SSAs that had one or more exceedances. Comparisons with SSAs with sample sizes less than 30 (indicated in figure by cross-hatching, and in figure legend by "n<30") were deemed unreliable and were therefore not included in the analysis. The 3 SSAs that were found to have the most number of significant comparisons are located east of the WTC. The numbers of exceedances for these three SSAs range from 2 to 9; their exceedance rates range from 0.021 to 0.060. The spatial pattern exhibited above is similar to the pattern of exceedance rates that is shown in Figure 3-13 however, 4 of the 7 SSAs with exceedance rates in the 4th quartile (Figure 3-13) were found to be significantly different from 5 or fewer of the other SSAs.

Table D-3a. *Test only* SSAs pairs having significant pairwise comparisons of exceedance rates.			
Significantly different means at the α =0.10 level			
10008002	10015011	10015011	10033001B
10008002	10039004	10015011	10033001A
10015011	10039001A	10033003B	10033001A
10015011	10033002B	10033002B	10033001A
Significantly different means at the α =0.05 level			
10007002	10015012	10015011	10039004
10008002	10021001	10015011	10039001B
10013002	10015011	10015011	10033003B
10013003	10015011	10015011	10317019A
Significantly different means at the α =0.01 level			
10007002	10015022	10015012	10027001
10008002	10013002	10015012	10039004
10008002	10013003	10015012	10039001B
10008002	10015022	10015012	10039001A
10008002	10021002	10015012	10033003B
10008002	10039001B	10015012	10033002B
10008002	10039001A	10015012	10033001B
10008002	10033003B	10015012	10033001A
10008002	10033002B	10015012	10317019A
10008002	10033001B	10015012	10317019C
10008002	10033001A	10015012	10317019D
10008002	10317019A	10015021	10015022
10008002	10317019C	10015022	10021001
10008002	10317019D	10015022	10021002
10013002	10015012	10015022	10025001
10013002	10015022	10015022	10027001
10013003	10015012	10015022	10039004
10013003	10015022	10015022	10039001B
10015011	10015012	10015022	10039001A
10015011	10015022	10015022	10033003B
10015011	10021002	10015022	10033002B
10015011	10317019C	10015022	10033001B
10015011	10317019D	10015022	10033001A
10015012	10015021	10015022	10317019A
10015012	10021001	10015022	10317019C
10015012	10021002	10015022	10317019D

[a]See figure D-3 for a map of the statistical summary areas (SSAs).

Table D-3b. Distribution of significant difference of means by MCBG, *Test Only*			
MCBG	Number of significant differences	MCBG	Number of significant differences
10007002	2	10021002	4
10008002	19	10025001	3
10008003	2	10027001	5
10009001	1	10033001A	5
10013002	4	10033002B	8
10013003	8	10033003B	5
10015011	5	10039001A	5
10015012	17	10039001B	6
10015021	17	10317019A	9
10015022	21	10317019C	8
10021001	4	10317019D	14

[a]See figure D-3 for a map of the statistical summary areas (SSAs).

Clean and Test Data

Description of sample. A total of 24,375 samples were collected, of which 17 have no geographic labels. The total number of samples with a value that exceeds the threshold level is 102. These sample asbestos measurements were aggregated by location into 45 statistical summary areas (SSAs) for lower Manhattan (south of Canal Street). One of these SSAs is the site that housed the WTC; no data were collected for this plus an additional 6 SSAs.

Initial data analysis. The rareness of exceedances suggests that these data may be described by a Poisson model. One feature of a Poisson random variable is that its mean, μ, and its variance are equal (equidispersion), a property frequently violated by real world data. "Failure of the Poisson assumption of equidispersion has similar qualitative consequences to failure of the assumption of homoskedasticity" associated with the Gaussian distribution (Cameron and Trivedi, 1998, p. 77). The standard way of accommodating overdispersion (the presence of more variation than is expected for a Poisson random variable) is by replacing a Poisson random variable with a negative binomial random variable—which can be viewed as a gamma mixture of Poisson random variables. In doing so, the distribution of counts is viewed as either (1) having missing variables for the mean specification, or (2) being dependent (i.e., the occurrence of an event increases the probability of further events occurring). The most popular implementation of the negative binomial probability model specifies the variance as being quadratic in the mean, or

$$\mu + \eta\mu^2 = (1 + \eta\mu)\mu,$$

with the dispersion parameter, η, to be estimated. The magnitude of η may be interpreted as follows (after Cameron and Trivedi, 1998, p. 79):

$\eta = 0$ implies no overdispersion;

$\eta \approx \frac{1}{\mu}$ implies a modest degree of overdispersion; and,

$\eta \geq \frac{2}{\mu}$ implies considerable overdispersion.

In other words, if $0 \leq \eta < \frac{0.5}{\mu}$, a spatial analyst may consider overdispersion detected in georeferenced data to be inconsequential, with little to be gained by replacing a Poisson with a negative binomial model specification. Meanwhile, recognizing that these exceedances are constrained by the number of samples collected suggests that these data may be described by a binomial model. Recoding counts of

exceedances to a binary (0-1) presence/absence measurement suggests that these data may be described by a logistic model. Simple estimation results for each of these four models appear in Table D-4.

Table D-4. Selected constant mean model estimation results for rates.

Model	intercept	equidispersion
Poisson for rates	-5.4756	NA
Negative binomial for rates	-5.2098	2.8692
Binomial	-5.4713	NA
Logistic	0.3185	NA

NOTE: rates were modeled by including the log of the number of cases as an offset variable.

One important implication from the tabulated results appearing in Table D-4 is that a Poisson model description of rates may suffer from a dramatic violation of the equidispersion assumption. The following evidence supports this claim:

$$\frac{2}{\hat{\mu}} = \frac{2}{2.68421} \approx 0.74510 << 2.8692.$$

In other words, the mean and variance may not be constant across the 38 SSAs.

Accounting for spatial autocorrelation. A conventional spatial autocorrelation analysis is hindered by two features of the collected data. One is the rareness of exceedance. In order to further explore spatial dependency in this context, average measures of asbestos also were analyzed. Both the rates and the average measures were transformed, using a logarithmic (i.e., Box-Cox 0 power) transformation with a translation parameter, to better conform to a bell-shaped curve (see Figure D-4). The other drawback is the absence of data for 7 SSAs. Because these areal units are dispersed across the study region, computing a Moran Coefficient (MC) becomes problematic.

MC scatterplots appear in Figure D-5. A conspicuous geographic pattern of positive spatial autocorrelation is apparent for the averages, and a possible positive spatial autocorrelation pattern may be present for the rates. Both patterns are corrupted by the presence of a number of zeroes.

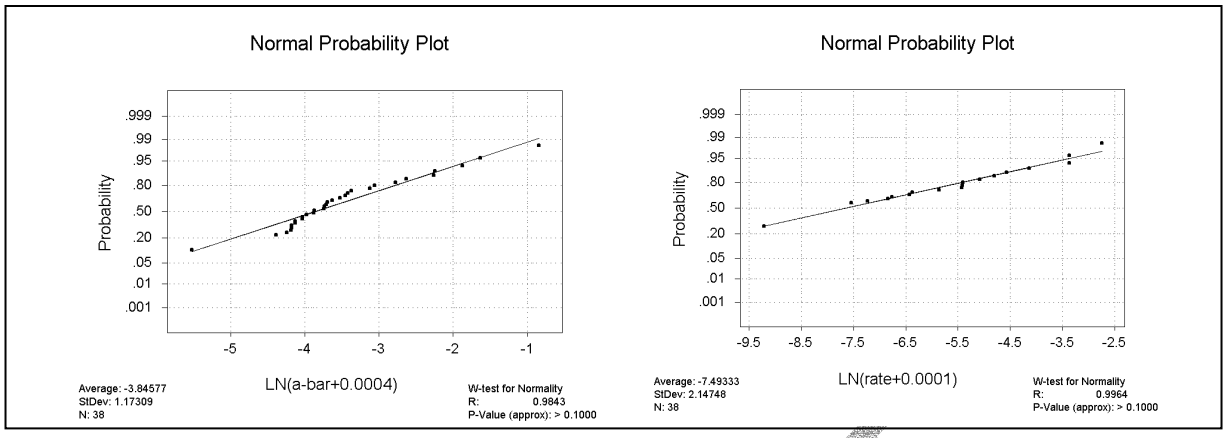

Figure D-4. Left: quantile plot for LN($\overline{\text{asbestos}}$ + 0.004). Right: quantile plot for LN(rate + 0.0001).

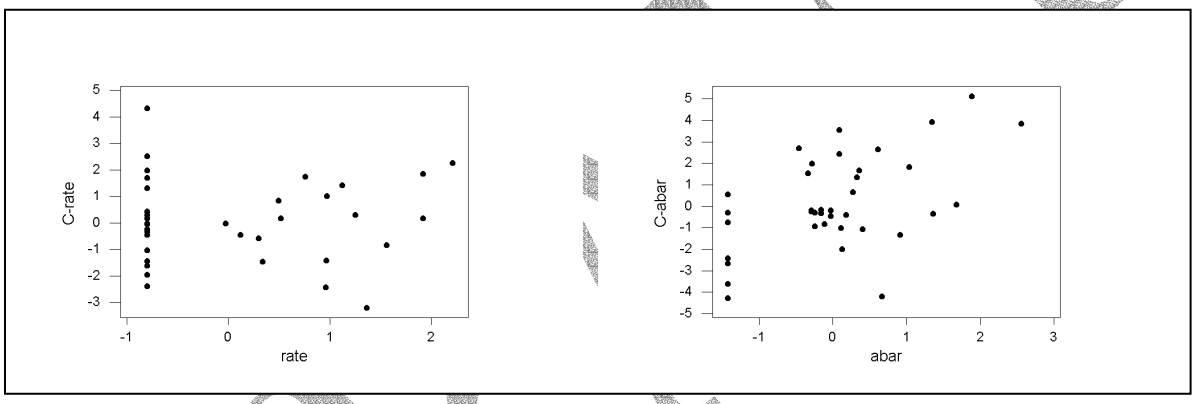

Figure D-5. Left: MC scatterplot for the rate of exceedance. Right: MC scatterplot for the average measure of asbestos.

Latent map patterns also can be assessed with eigenvectors derivable from a MC. Here five of the 11eigenvectors (E_2, E_3, E_8, E_{17}, E_{22}; these were selected using the stepwise options for PROC LOGISTIC in SAS, and SWPOIS in STATA) denoting consequential positive spatial autocorrelation help describe the geographic distribution of exceedance rates for the Poisson and binomial models. The negative binomial model failed to be estimable, yielding a negative maximum likelihood estimate for dispersion; but, the deviance measure for the estimated Poisson model is 1.36, suggesting a lack of serious overdispersion. One eigenvector (E_{10}) relates to the logistic version of the variable. Maps of three of the five synthetic geographic variables appear in Figure D-6.

The MC values for the five eigenvectors range from 0.38 to 0.93. Estimation results that include these eigenvectors in the specifications of each of the four models appearing in Table D-4 are reported in Table D-5. The Poisson model with an assumption of equidispersion appears to be reasonable here. This

specification accounts for roughly 40% of the variation in the geographic distribution of rates. The binomial model specification renders very similar results. The logistic model description seems inappropriate.

One important finding that can be gleaned from Table D-2 is even detected modest overdispersion accompanying the Poisson model description largely is attributable to latent spatial autocorrelation.

Table D-5. Selected model estimation results for rates when spatial dependence is included.

Variable	Poisson model	Negative binomial model	Binomial model	Logistic model
intercept	-5.9383	Failed to be estimable	-5.9347	0.2773
equidispersion	NA		NA	NA
E_2	-4.2422		-4.2812	NA
E_3	4.4056		4.4575	NA
E_8	-5.4424		-5.5115	NA
E_{10}	NA		NA	-4.4191
E_{17}	-2.2640		-2.2961	NA
E_{22}	4.2200		4.2615	NA

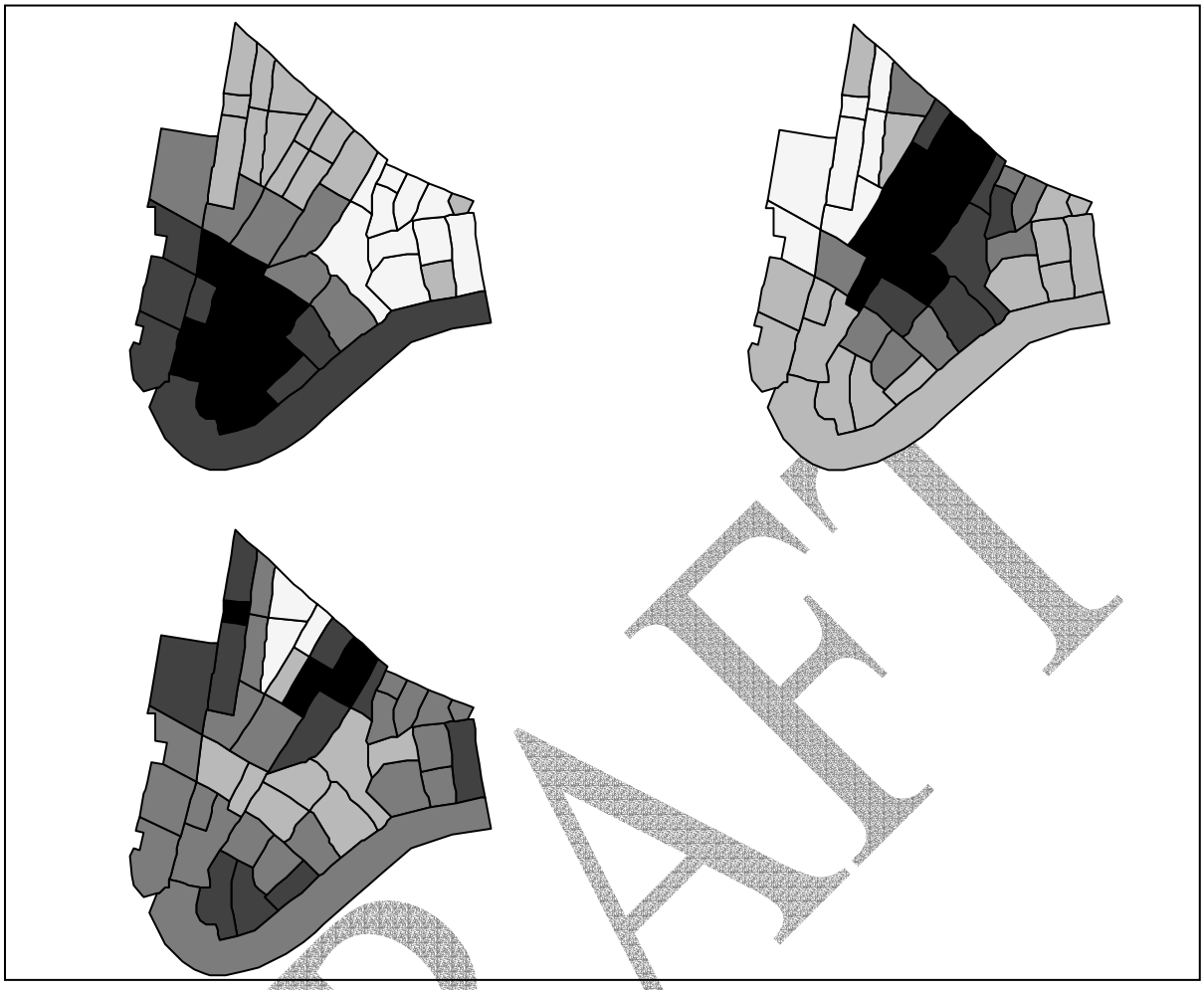

Figure D-6. Geographic distribution of relevant eigenvectors. Top left: E_2. Top right: E_3. Bottom left: E_{17}.

Pairwise comparisons between SSAs. Pairwise comparisons of SSA asbestos exceedance sampling results were made to assess whether or not statistically significant differences exist. Aggregate sample sizes less than 30 are considered too unreliable, and were not included in this assessment. The outcome of this sample size restriction is 32 SSAs with a sufficient number of samples, allowing ($31 \times 30/2 =$) 465 pairwise comparison.

Differences of exceedance rates. The estimated spatially filtered Poisson model produces sample mean estimates for uncorrelated Poisson variables. These models include LN(# of cases) as an offset variable. Therefore, dividing both sides of the estimated equation for $\hat{\mu}_i$ (i.e., the mean rate for areal unit i) by the

corresponding number of samples yields the set of estimated rates, assuming an underlying Poisson process, of $\frac{\hat{\mu}_i}{n_i}$, i=1, 2, …, 31. The accompanying set of null hypotheses becomes

$$H_o: \frac{\mu_i}{n_i} - \frac{\mu_j}{n_j} = 0 \text{ , } i \neq j, \text{ i=1, 2, …, 31 and j=i+1, i+2, …, 31 .}$$

The estimated standard error for this difference of rates test is given by $\sqrt{\frac{\hat{\mu}_i}{n_i^2} + \frac{\hat{\mu}_j}{n_j^2}}$.

A simulation experiment involving 50,000 difference of means replications (total=496×50,000) was conducted using the spatially filtered Poisson model estimation results. The simulated Poisson random variable, Y, then was used in a bivariate linear regression analysis, which yielded

$$\frac{\hat{\mu}_i}{n_i} - \frac{\hat{\mu}_j}{n_j} = -0.00000 + 0.99955 \overline{\frac{y_i}{n_i} - \frac{y_j}{n_j}} + e \text{ , } R^2 = 1.00, \text{ and}$$

$$\sqrt{\frac{\hat{\mu}_i}{n_i^2} + \frac{\hat{\mu}_j}{n_j^2}} = 0.00001 + 0.99873 \, s_{\frac{y_i}{n_i} - \frac{y_j}{n_j}} + e \text{ , } R^2 = 1.00.$$

In the second case, the intercept is significantly different from 0, and in both cases the slope is significantly different from 1. These may well be size effect results, since substantively both intercepts effectively are zero, and both slopes effectively are 1.

The model-based mean estimates range from roughly 0.10 to 32.60, implying that at least some of the difference of rates sampling distributions should be non-normal. Each simulated dataset was subjected to a diagnostic Kolmogorov-Smirnov goodness-of-fit test for a normal distribution, producing test statistics ranging from roughly 0.01 to 0.35. In other words, the simulated sampling distributions fail to conform to normal distributions. Consequently, the pairwise difference of rates assessments are based upon a Hope-type nonparametric simulation analysis, involving 99,999 replications coupled with each observed difference. The simulated distribution is based on a pair of Poisson random variables, each with the same mean of $\frac{n_1\mu_2 + n_2\mu_1}{2n_1n_2}$, which yields a null hypothesis difference of 0 and the correct theoretical variance

of $\frac{\mu_1}{n_1} + \frac{\mu_2}{n_2}$. Because a two-tailed test is employed here, an observed rank of 1 or 100,000 results in a rejection of the null hypothesis for α=0.01, an observed rank of 2-5 or 99,996-99,999 results in a rejection of the null hypothesis for α=0.05, and an observed rank of 6-11 or 99,990-99,995 results in a rejection of the null hypothesis for α=0.10. Based on these criteria, six pairs of exceedance rates are significantly different at the 10% level, 14 pairs are significantly different at the 5% level, and 122 pairs are significantly different at the 1% level. Basically, roughly 33% of the extreme MCGB mean pairs tend to be significantly different. These differences arise from four clusters of mean sizes. The first is the extreme MCBG mean of nearly 33 (MCBG 10015022). The second is the third largest mean of approximately 16 (MCBG 10015012). The third is the somewhat small mean of 0.64 (MCBG 10016004) which more than likely is being amplified by its small sample size of 32. The remaining 28 MCGBs form a set whose sample-size-weighted absolute differences of means range from nearly 0 to almost 0.1. Primarily, significant differences are between the extremes within this group (see Tables D-6a and 6b, and Figure D-7).

These results need to be moderated by keeping in mind that the estimated Poisson model accounts for only about 50% of the variance in the observed exceedances.

Table D-6a. SSAs pairs having significant pairwise comparisons of rates[a].					
Significantly different means at the α = 0.10 level					
10009002	10027001	10015011	10033002A	10027001	10317019C
10009002	10033003B	10015021	10317019C	10039004	10317019C
Significantly different means at the α = 0.05 level					
10007002	10009002	10013003	10027001	10021001	10317019C
10008002	10021001	10021001	10021002	10027001	10039001A
10009001	10027001	10021001	10025001	10033003B	10317019C
10009002	10015021	10021001	10039001B	10039004	10033003B
10013003	10021001	10021001	10039004		
Significantly different means at the α = 0.01 level					
10007002	10009001	10015011	10021001	10015021	10021002
10007002	10013002	10015011	10021002	10015021	10025001
10007002	10015011	10015011	10025001	10015021	10039001B
10007002	10015012	10015011	10027001	10015021	10039004
10007002	10015022	10015011	10029002	10015021	10317019A
10007002	10021002	10015011	10031001	10015021	10317019D
10007002	10039004	10015011	10033001A	10015022	10016004
10007002	10317019D	10015011	10033001B	10015022	10021001
10008002	10015011	10015011	10033002B	10015022	10021002
10008002	10015012	10015011	10033003A	10015022	10025001
10008002	10015022	10015011	10033003B	10015022	10027001
10008002	10027001	10015011	10039001A	10015022	10029002
10008003	10015012	10015011	10039001B	10015022	10031001
10008003	10015022	10015011	10039003	10015022	10033001A
10009001	10015011	10015011	10039004	10015022	10033001B
10009001	10015012	10015011	10317019A	10015022	10033002A
10009001	10015021	10015011	10317019C	10015022	10033002B
10009001	10015022	10015011	10317019D	10015022	10033003A
10009001	10021001	10015012	10015021	10015022	10033003B
10009001	10033003B	10015012	10015022	10015022	10039001A
10009001	10317019C	10015012	10016004	10015022	10039001B
10009002	10015011	10015012	10021001	10015022	10039003
10009002	10015012	10015012	10021002	10015022	10039004
10009002	10015022	10015012	10025001	10015022	10317019A
10009002	10021001	10015012	10027001	10015022	10317019C
10009002	10317019C	10015012	10029002	10015022	10317019D
10013002	10015011	10015012	10031001	10021001	10039001A
10013002	10015012	10015012	10033001A	10021001	10317019A
10013002	10015021	10015012	10033001B	10021001	10317019D
10013002	10015022	10015012	10033002A	10021002	10027001
10013002	10021001	10015012	10033002B	10021002	10033003B
10013002	10027001	10015012	10033003A	10025001	10027001
10013002	10033003B	10015012	10033003B	10027001	10039001B
10013002	10317019C	10015012	10039001A	10027001	10039004
10013003	10015011	10015012	10039001B	10027001	10317019A
10013003	10015012	10015012	10039003	10027001	10317019D
10013003	10015021	10015012	10039004	10033003B	10317019A
10013003	10015022	10015012	10317019A	10033003B	10317019D
10015011	10015012	10015012	10317019C	10039001B	10033003B
10015011	10015021	10015012	10317019D	10317019C	10317019D
10015011	10015022	10015021	10015022		

[a]See Figure D-7 for a map of the statistical summary areas (SSAs).

Table D-6b. Distribution of significant difference of means by MCBG, Clean & Test Data

MCBG	Number of Significant differences	MCBG	Number of Significant differences
10007002	9	10029002	3
10008002	5	10031001	3
10008003	2	10033001A	3
10009001	9	10033001B	3
10009002	9	10033002A	3
10013002	9	10033002B	3
10013003	6	10033003A	3
10015011	28	10033003B	12
10015012	30	10039001A	5
10015021	14	10039001B	7
10015022	30	10039003	3
10016004	2	10039004	9
10021001	16	10317019A	7
10021002	8	10317019C	12
10025001	6	10317019D	9
10027001	16		

[a]See Figure D-7 for a map of the statistical summary areas (SSAs).

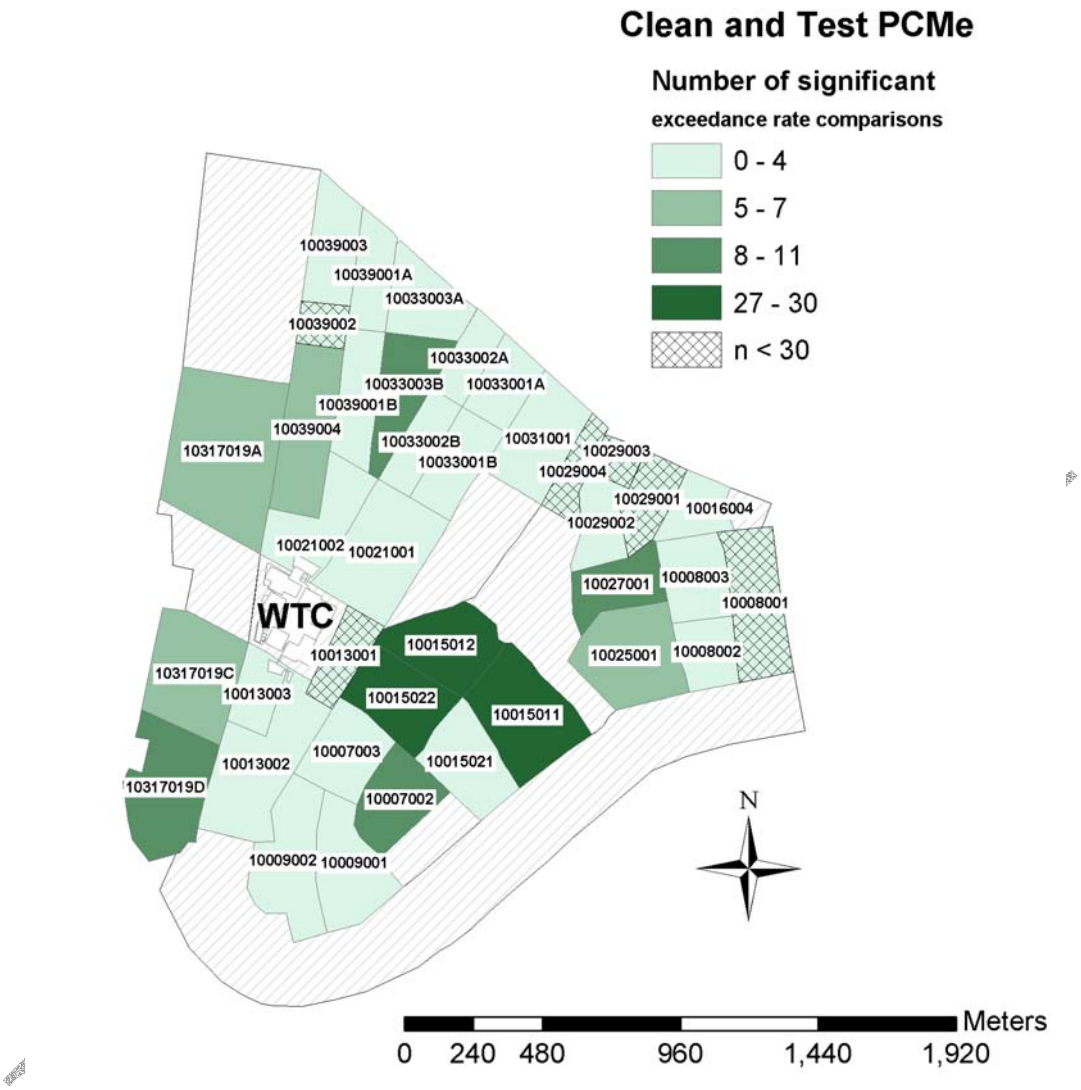

Figure D-7. Significant differences between estimated exceedance rates for *clean and test* data, with Statistical Summary Areas labeled. Estimates are based on the spatially-filtered Poisson model (see Section 3.2.3.2 and Appendix D for details). The number of significant pairwise comparisons at an experiment-wise $\alpha = 0.01$ (with a Bonferroni adjustment) are shown for SSAs that had one or more exceedances. Comparisons with SSAs with sample sizes less than 30 (indicated in figure by cross-hatching, and in figure legend by "n<30") were deemed unreliable and were therefore not included in the analysis. Three of the SSAs that were found to have the most number of significant comparisons are located east of the WTC. The numbers of exceedances for these three SSAs range from 17 to 32; their exceedance rates range from 0.006 to 0.059. The spatial pattern exhibited above is similar to the pattern of exceedance rates that is shown in Figure 3–14 however, 3 of the 9 SSAs with exceedance rates in the 4th quartile (Figure 3–14) were found to be significantly different from 4 or fewer of the other SSAs.

Clean and Test Data Subset

Description of sample. When sampling results for five intensively sampled buildings are removed from the *clean and test* dataset, a total of 17,905 samples remain, of which 17 have no geographic labels. The total number of samples with a value that exceeds the threshold level is 92. These sample asbestos measurements were aggregated by location into 45 modified census block groups (SSAs) for lower Manhattan. One of these SSAs is the site that housed the WTC; the modified database contains no data for this plus an additional 7 SSAs.

Initial data analysis. Simple estimation results for each of the four models (i.e., Poisson, negative binomial, binomial and logistic) that parallel those for the complete dataset appear in Table D-7. These results are very similar to those obtained with the complete dataset, too.

Table D-7. Selected constant mean model estimation results for rates.

Model	intercept	equidispersion
Poisson for rates	-5.2701	NA
Negative binomial for rates	-5.1409	3.1819
Binomial	-5.2649	NA
Logistic	0.4964	NA

NOTE: rates were modeled by including the log of the number of cases as an offset variable.

Accounting for spatial autocorrelation. Identified prominent latent map patterns also are very similar (E_3, E_8, and E_{17} are common to the rates models; and again were selected using the stepwise options for PROC LOGISTIC in SAS, and SWPOIS in STATA). One of the eigenvectors identified with the complete dataset disappears here (E_2). One model difference now is that the binomial model links to eigenvector E_{22}, whereas the Poisson model links to eigenvector E_{27}. The negative binomial model yielded a dispersion parameter estimate of 0 here, making it indistinguishable from a Poisson model. As before, the same single eigenvector (E_{10}) relates to the logistic version of the variable.

The Poisson model with an assumption of equidispersion appears to be reasonable here. This specification accounts for roughly 40% of the variation in the geographic distribution of rates.

Table D-8. Selected model estimation results for rates when spatial dependence is included.				
Variable	Poisson model	Negative binomial model	Binomial model	Logistic model
intercept	-5.8875	Failed to be estimable	-5.8827	0.5075
equidispersion	NA		NA	NA
E_1	2.2292		2.2311	NA
E_3	4.1741		4.2493	NA
E_8	-3.4133		-3.4547	NA
E_{10}	NA		NA	-5.3629
E_{17}	-2.8619		-2.8848	NA
E_{22}	NA		NA	NA
E_{27}	-3.3557		-3.4310	NA